Materials Data Infrastructure and Ontology Development for Exemplary Reference Dataset

Angelika Gedsun, Yusra Shakeel, Sirieam Hunke, Luis A. Ávila C., Mariano Forti, Thomas Hammerschmidt, Jürgen Olbricht, Rainer Stotzka, Tilmann Hickel, Birgit Skrotzki

Introduction

Reference data (RD) is reliable data that are of utmost importance for sustainable materials development. Researchers rely on such RD especially for the validation of research results in domains or for materials where data is scarce. Within NFDI-MatWerk, we develop an infrastructure to distribute and validate reference datasets on the example of creep data. With our materials data infrastructure (MDI), we aim to understand and structure data along with its metadata to make it findable, accessible, interoperable, and reusable (FAIR).

In this poster, we present a so-called **data journey** (middle) depicting the processes and stages within a data lifecycle along with selected elements that are the basis to build our MDI for RD. A community-agreed **data schema** enables machine actionable representation as **FAIR Digital Object** (FDO). The main concepts of the dataset are semantically described and represented in the form of the Reference Dataset **Ontology** for Creep data (RDOC), and integrated into the MSE Knowledge Graph.

Our Goal

Conceptualize & implement a digital infrastructure for
- Shaping, storing, sharing
- Search, discovery, and usage of reference data

Key Takeaways

The infrastructure enables the researcher to:
- Validate a dataset as a reference dataset → Metadata schema
- Search/find data → FDO and instantiated ontology

Data Schema

- List of possible and use-case-related minimum requirements
- Usage of a defined **vocabulary**
- RDM Tool to enhance documentation practices
- Serves as a template for other methods
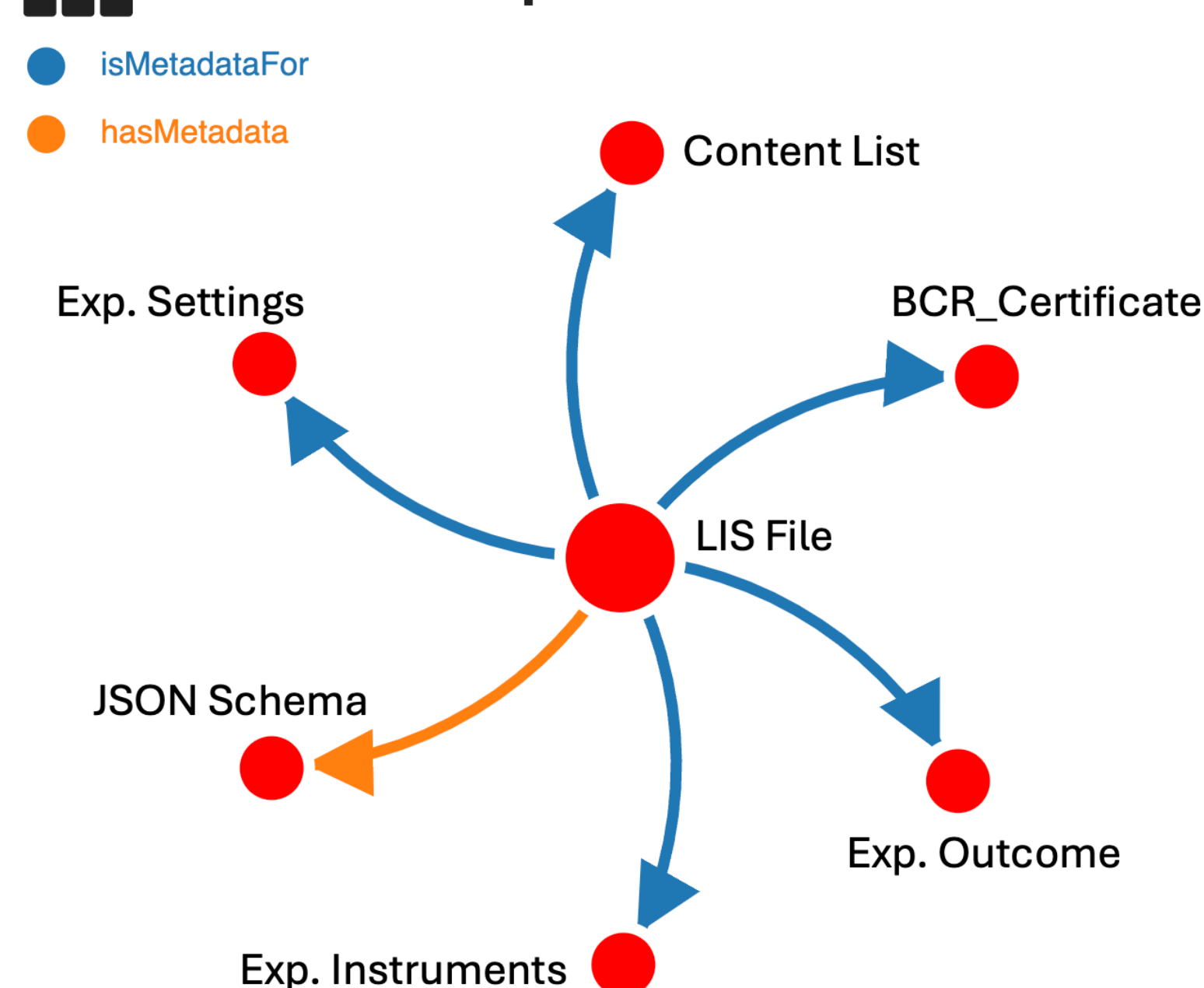- Basis for shaping, sharing, and structuring the data
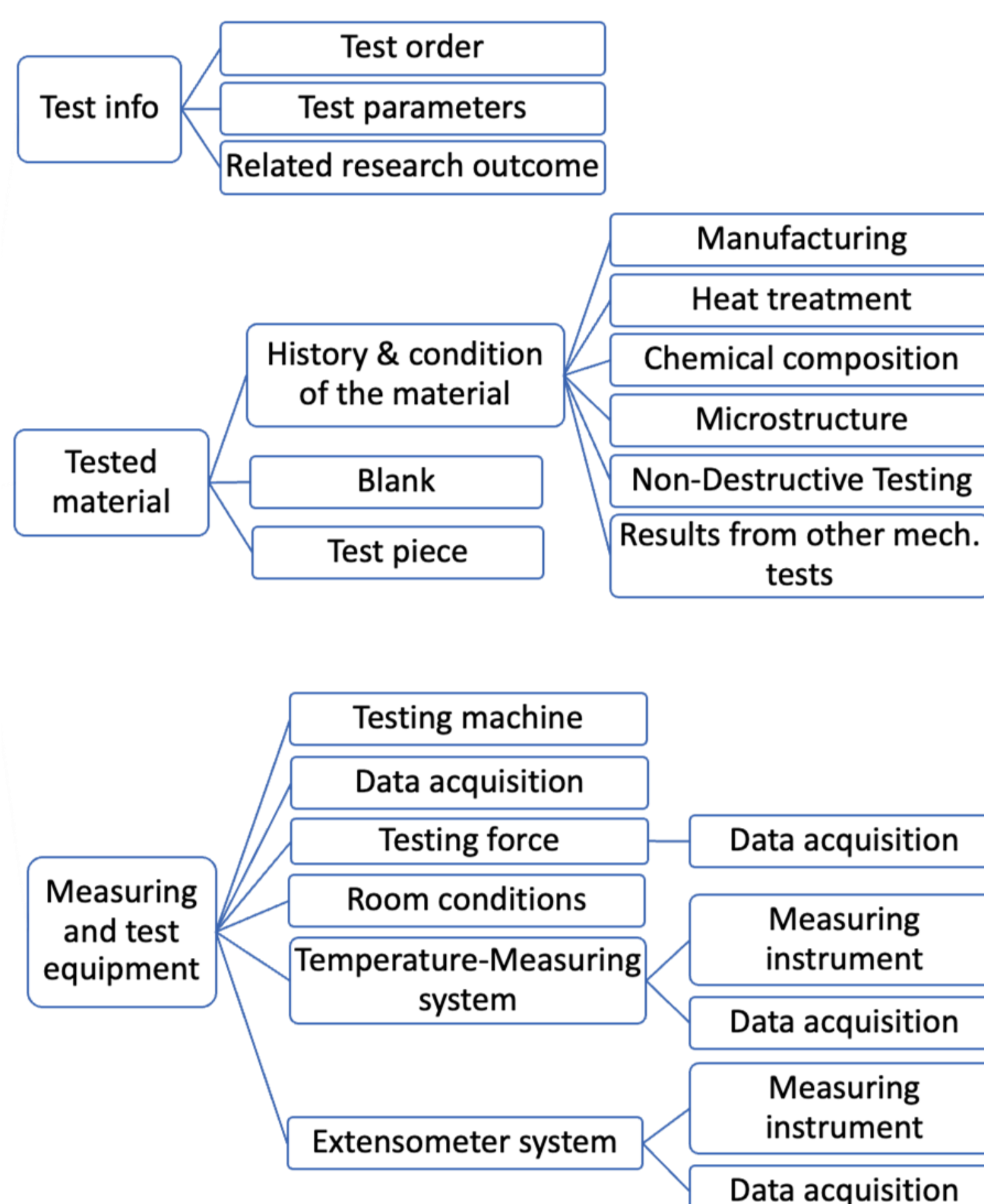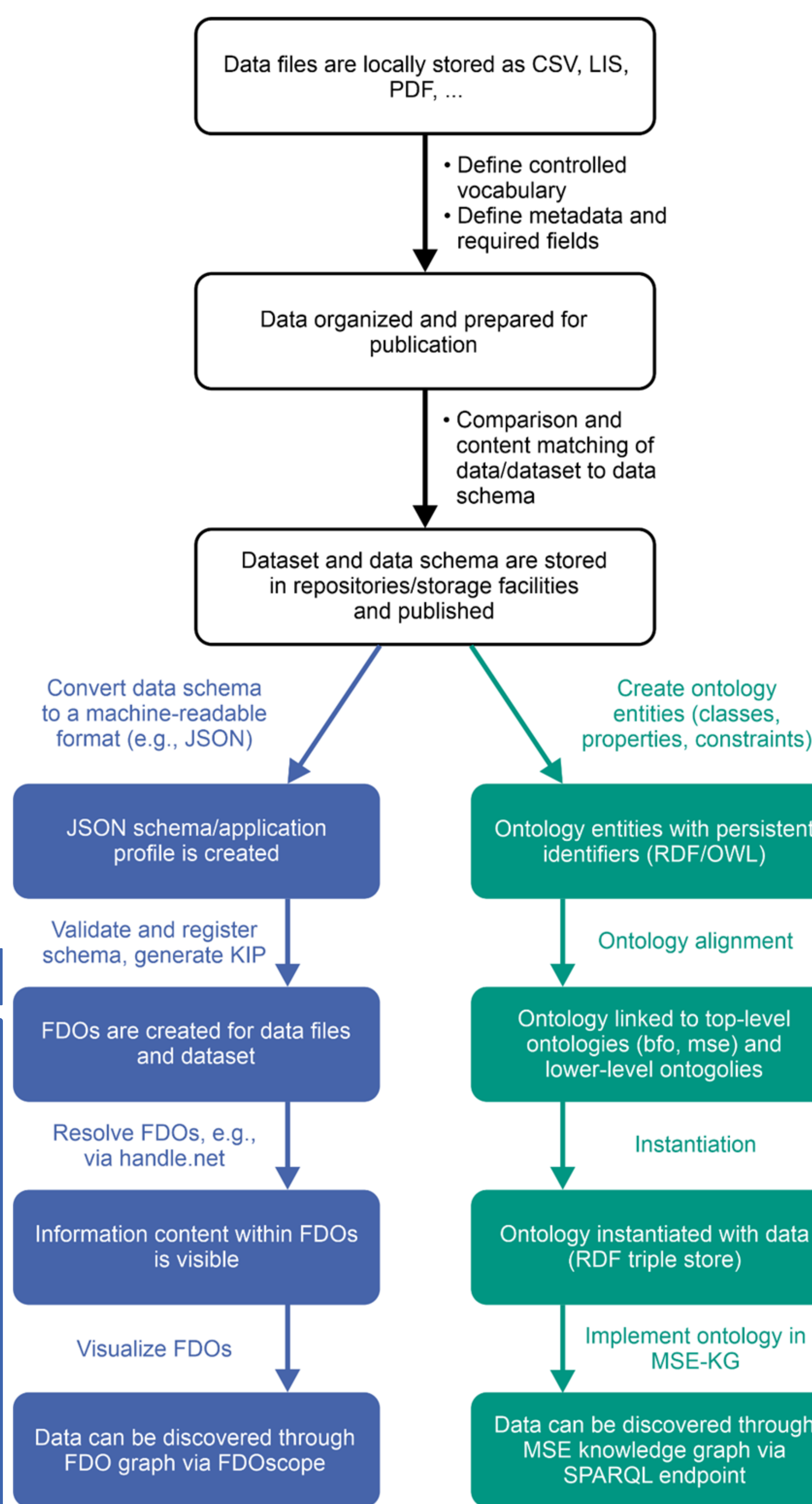

Data Discovery and Usage

FAIR Digital Object (FDO)
- Globally accepted representation of research data that makes FAIR data not only machine readable but also machine actionable across scientific disciplines
- Allows a common view on all data artifacts independent of their scientific contents and metadata structures

Schematic View of the Data Journey

Data files are locally stored as CSV, LIS, PDF, ...

- Define controlled vocabulary
- Define metadata and required fields

Data organized and prepared for publication

- Comparison and content matching of data/dataset to data schema

Dataset and data schema are stored in repositories/storage facilities and published

Convert data schema to a machine-readable format (e.g., JSON)

JSON schema/application profile is created

Validate and register schema, generate KIP

FDOs are created for data files and dataset

Resolve FDOs, e.g., via handle.net

Information content within FDOs is visible

Visualize FDOs

Data can be discovered through FDO graph via FDOscope

Create ontology entities (classes, properties, constraints)

Ontology entities with persistent identifiers (RDF/OWL)

Ontology alignment

Ontology linked to top-level ontologies (bfo, mse) and lower-level ontologies

Instantiation

Ontology instantiated with data (RDF triple store)

Implement ontology in MSE-KG

Data can be discovered through MSE knowledge graph via SPARQL endpoint

Ontology Development

Concepts for a universal Reference Dataset

Concepts for a Reference Dataset for Creep Testing

→ Instantiate ontology
→ Integrate into MSE-KG
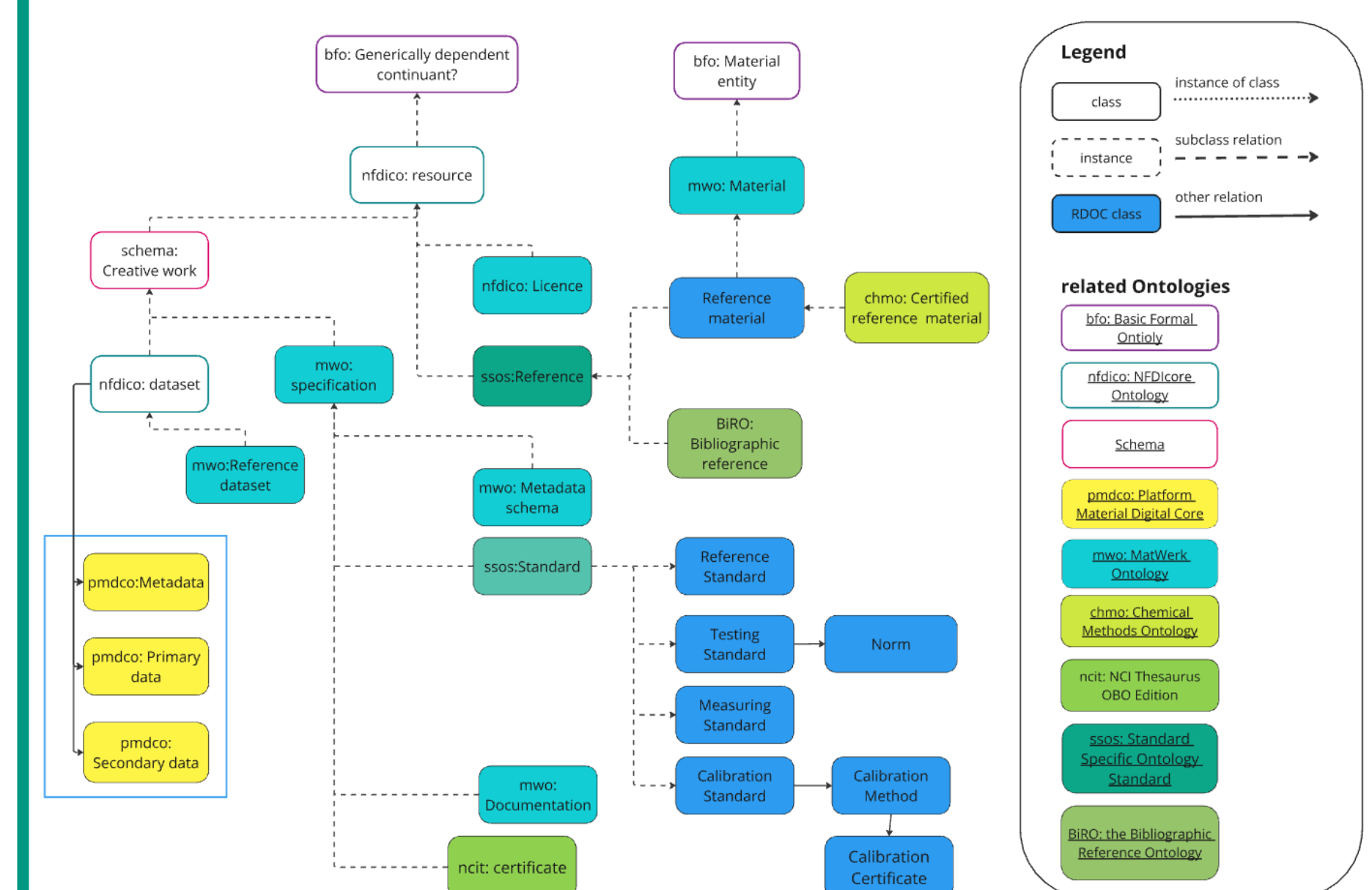→ Link to other resources in MSE community
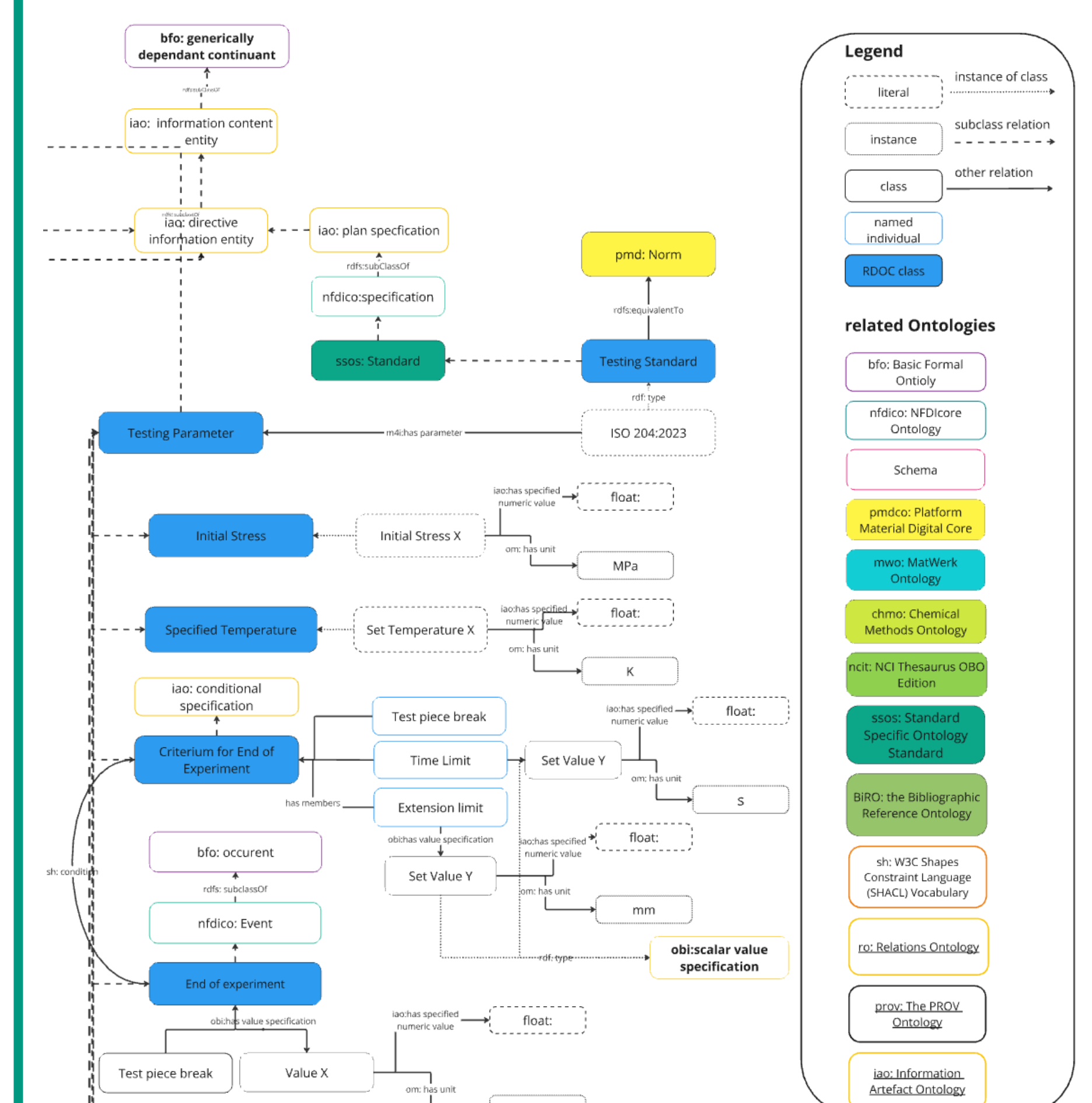
Possible usages of reference data

- Calibration/verification of measurement devices, procedures or algorithms
- Comparison of data for interpretation of individual measurement results
- Input for machine learning-based data analytics and for computational materials science (digital twin)
- Best practice examples for measurement and documentation procedures

angelika.gedsun@imtek.uni-freiburg.de
yusra.shakeel@kit.edu